**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF NEW JERSEY**

JESSICA ROBINSON, STACEY JENNINGS, and NICOLE GIBSON, *individually and on behalf of all others similarly situated,* *Plaintiffs,* v. JACKSON HEWITT INC., and TAX SERVICES OF AMERICA, INC., *Defendants*.	Civil Action No.: 2:19-cv-9066(SDW)(ESK) **STIPULATED ORDER REGARDING ESI PROTOCOL AND FORMAT OF PRODUCTIONS**

THIS MATTER came before the Court on the joint request of Plaintiffs and Defendants for entry of an Order, and for good cause shown, it is hereby **ORDERED** that the format of documents, including electronically stored information ("ESI") and paper documents, produced in this case shall be in accordance with the following protocol.

I.   General Provisions

      A.   Scope. This Order is intended to give effect to Fed. R. Civ. P. 1, *i.e.,* "to secure the just, speedy and inexpensive determination of every action and proceeding." An important purpose of this protocol is to reduce litigation costs and expedite the litigation without compromising any party's position on the underlying merits. This purpose shall be referenced in resolving any disputes that arise as to the interpretation of any provision of this Order.

      B.   Cooperation. The parties will cooperate on issues relating to discovery of ESI and other documents. The parties recognize that collaborative conduct in the course of discovery reduces litigation costs and delay. The parties will engage in informal discussions about the discovery of ESI throughout the litigation as necessary.

Counsel will be knowledgeable about their respective clients' information technology infrastructure, including the manner in which potentially relevant data is stored and retrieved. When appropriate, counsel may engage the assistance of someone knowledgeable about a party's electronic systems, as necessary.

C.   Discovery Confidentiality Order. The parties agree that this Stipulated Order is subject to the previously entered Discovery Confidentiality Order entered in this action (Docket No. 86) (the "Confidentiality Order"), and they do not intend to waive or modify any portion of that order.

D.   Avoidance of Unnecessary Waste. No party will produce ESI to another party in a manner that unnecessarily diminishes the ability of the ESI to be searched or that unnecessarily results in the loss of the ESI's structure, categorization, or utility. The purpose of this provision is to avoid unnecessary expense and duplicative effort by causing a party to recreate structure, organization, or searchable format where it already exists.

E.   Not reasonably accessible ("NRA") data sources. A producing party retains the right to argue that certain sources of ESI are not reasonably accessible because of undue burden or cost, and further retains the right to seek cost shifting, as appropriate, in the event a requesting party demands and shows good cause supporting production from sources that the producing party deems not reasonably accessible. In acknowledging this reservation of rights neither party concedes as an initial matter that any source is not reasonably accessible. To the extent a party believes that responsive documents are not reasonably accessible, it shall immediately identify, by category or type, the source(s) containing potentially

2

responsive information with enough detail to enable the requesting party to evaluate the burdens and costs of providing the discovery and the likelihood of finding responsive information from the identified source(s). Absent a showing of good cause or written agreement of the parties, there is no need to preserve or collect ESI from the following sources, which are deemed by agreement to be NRA in this particular litigation:

1.   telephonic voice mails;

2.   backup tapes that are duplicative of data preserved in an equally or more accessible format;

3.   "slack," "fragmented," or "unallocated" data on hard drives, provided that this Order is not intended to restrict otherwise permissible discovery of information about the deletion of such ESI;

4.   random access memory (RAM) or other ephemeral data;

5.   on-line access data (temporary internet files, history, cache, cookies, etc.);

6.   disaster recovery backup media that is duplicative of data preserved in an equally or more accessible format;

7.   data in metadata fields that are frequently updated automatically, such as last-opened dates;

8.   other forms of ESI the preservation of which requires extraordinary affirmative measures that are not utilized in the ordinary course of business.

F.   <u>Date Range for Search and Review of Documents.</u> Each party shall search for and review ESI and documents accessed, modified, created, sent, or received from December 20, 2014 to December 20, 2018, without precluding a party from seeking

823032.1

specific discovery outside of this date range. The date limitation in this section shall not apply to the parties' production of Structured Data as defined and described in § II.A.4.(l) below.

G.   <u>Custodians and Search Terms.</u> Before implementing search terms, the producing party will provide to the receiving party a list of proposed search terms, custodians and non-custodial sources the party proposes to search and the parties shall confer in good faith to identify and agree on appropriate search terms, custodians, and sources. Before implementing search terms, the producing party will provide a search term hit list or hit report that should include the number of documents that hit on each term, the number of unique documents that hit on each term (documents that hit a particular term and no other term on the list) and the total number of documents that would be returned by using the proposed search term list (including families). The parties will use the report to meet and confer regarding any potential modifications to terms. Nothing in this Order shall prevent any party from conducting a manual review of documents for relevance, privilege, and confidentiality after application of search terms. To the extent that documents responsive to requests served under Fed. R. Civ. P. 34 can be identified and produced without the use of search terms, custodians, or date limitations, the parties shall produce those responsive documents without listed limitations or search criteria

H.   <u>Resolution of Disputes.</u> The parties shall confer in good faith to resolve any disputes arising under this Order before submitting such disputes to the Court.

II.   <u>Production Format Specifications</u>

A.   General Specifications

1.   Produce in Native and TIFF Format. The parties will produce electronic

documents in TIFF Format as described below, other than the structured

data addressed in Paragraph II.A.4(l), and   Excel and Powerpoint

documents, which will be produced in native format.[1] Production of

structured data is subject to the parties' future negotiations regarding the

format for production of structured data as contemplated in Paragraph

II.A.4(l). Each page of each paper document for production shall be

electronically scanned to create a single-page TIFF image file. The required

specifications for TIFF image files are contained in Appendix 1. The

required specifications for native files are contained in Appendix 5.

Without waiving their respective rights to request cost-shifting in

appropriate circumstances pursuant to Fed. R. Civ. P. 26, each party shall

bear the cost associated with any TIFF conversion performed in connection

with producing its own documents. In acknowledging this reservation of

rights neither party concedes as an initial matter that any source is not

reasonably accessible. Each party shall bear its own costs for TIFF

conversion of its own documents.

2.   Provide Additional Standard Files. In addition to the TIFF image files

referenced above, the parties shall supply with each production: (a) a data

load file, (b) text files, and (c) an image load file corresponding to the

---

[1] Electronic documents in other formats not easily navigated or organized in TIFF Format may
also need to be produced in native format, which the parties will negotiate at the appropriate
time.

823032.1

produced TIFF image files as set forth below.

    a.   <u>Data Load File</u>. The data load file shall contain certain metadata about each produced document. The required specifications for the data load files are contained in Appendix 2.

    b.   <u>Text Files</u>. A single text file shall be provided for each respective document, not one text file per page, containing the readable text within that document obtained using either a standard OCR process (for scanned paper documents or redacted ESI only) or a standard text extraction process (for all other ESI). The required specifications for the text files are contained in Appendix 3.

    c.   <u>Image Load File</u>. The produced TIFF images shall be provided with a standard Opticon image load file that can be loaded into standard litigation support software. The required specifications for the image load file are contained in Appendix 4.

3.   <u>Unitizing of Paper Documents</u>. In scanning paper documents, the unitization of the document and any attachments shall be maintained as it existed in the original when creating the image file. Distinct documents shall not be merged into a single record, and single documents shall not be split into multiple records (i.e. paper documents should be logically unitized). Parent-child relationships (the association between an attachment and its parent document) shall be preserved. Documents will be unitized at the lowest possible level and attachment information preserved. For example, if a folder contains two documents, the folder and each document

will constitute a separate document, but the relationship among the documents in the folder should be reflected in proper coding of the fields in the data load file.

4.   Additional ESI Processing Specifications.

a.   De-duplication. "Duplicate ESI" means files that are exact duplicates based on the files' MD5 or SHA-1 hash values at the family level. The producing party need produce only a single copy of responsive Duplicate ESI. A producing party shall take reasonable steps to de-duplicate ESI globally (i.e. both within a particular custodian's files and across all custodians). Entire document families may constitute Duplicate ESI. De-duplication shall not break apart families. When the same Duplicate ESI exists in the files of multiple custodians, those persons shall be listed in the AllCustodian field identified in Appendix 2. To the extent that deduplication is used, the parties expressly agree that a document produced from one custodian's file but not produced from another custodian's file as a result of deduplication will nonetheless be deemed as if produced from that other custodian's file for purposes of deposition, interrogatory, request to admit and/or trial.

b.   Hidden Text or Data. The TIFF images of all files, including without limitation Microsoft Word files (*.doc and *.docx) and Powerpoint files (*.ppt), will be created with any tracked changes, comments, speaker's notes, or any other hidden data showing in

7

such images.

d.   Spreadsheet or Worksheet Files.   Spreadsheet files, including

without limitation Microsoft Excel files (*.xls or *.xlsx), will be

supplied in their native format.   A link to the native file will be

provided in the NATIVEPATH field.   A placeholder TIFF image

endorsed with PRODBEG will be supplied.

e.   Parent-Child Relationships.   Parent-child relationships (e.g. the

associations between emails and their attachments) will be

preserved.   Email and other ESI attachments will be produced as

independent files immediately following the parent email or ESI

record.

f.   Dynamic Fields. Files containing dynamic fields such as file names,

dates, and times will be produced showing the field code (e.g.

"[FILENAME]" or "[AUTODATE]"), rather than the values for

such fields existing at the time the file is processed.

g.   English Language. To the extent any data exists in more than one

language, the data will be produced in English, if available. If no

English version of a file is available, the producing party shall not

have an obligation to produce an English translation of the data.

h.   Embedded Objects.   Some Microsoft Office and .RTF files may

contain embedded objects.   Subject to claims of privilege or

immunity, as applicable, objects with such file types shall be

extracted as separate files and shall be produced as attachments to

8

the file in which they were embedded.

i.   Compressed Files. Compressed file types (e.g. .CAB, .GZ, .TAR. .ZIP) shall be decompressed in a reiterative manner to ensure that a compressed file within a compressed file is decompressed into the lowest possible compression resulting in individual files.

j.   Encrypted Files. The producing party will take reasonable steps to unencrypt any discoverable ESI that exists in encrypted format (e.g. password-protected) and that can be reasonably unencrypted. Files that cannot be unencrypted should nonetheless be produced in their original format. If a producing party cannot unencrypt discoverable ESI that exists in encrypted format, the parties agree to meet and confer regarding how such information should be handled.

k.   Email Threading. Email threads are email communications that contain prior email communications. A most inclusive email thread is one that contains all of the prior or lesser-included emails, including attachments, for that branch of the email thread. The producing party may remove from production the wholly-included, prior-in-time email communications. Additionally, the producing party may list on any required privilege log only the most inclusive email thread.

l.   Structured Data. The parties shall negotiate separately regarding the production of structured database files. When it is impractical to produce an entire database file, the parties will collaborate to

9

construct appropriate queries and reports necessary to extract discoverable ESI in a file format most consistent with the specifications of this agreement. The time frame applicable to structured data shall be December 20, 2014 to present.

    m.   <u>Multimedia Files</u>. A producing party will identify the existence of potentially discoverable ESI that is comprised of media such as audio files, video files, or photographic images. Any such files will be uniquely identified, and the parties shall negotiate separately regarding the production of multimedia files such as audio and video files, if any.

B.   <u>Redactions</u>. The producing party may redact privileged material from any TIFF image or metadata field. Nothing in this Order shall preclude a party from filing a motion for protective order or from seeking an *in camera* review for requested redactions to a specific document on grounds other than privilege. Redactions and disputes regarding redactions will be handled in accordance with the Confidentiality Order and rules of this Court. Each redaction shall be indicated clearly on the redacted document. OCR text generated from the redacted production image(s) will be provided in place of the original extracted text associated with a redacted document.

C.   <u>Family Groups</u>. A document and all other documents in its attachment range constitute a family group. If any document in a family contains a search term, all documents in that family shall be deemed to be responsive. Documents within the same family as produced documents may be withheld as privileged and placed on

<center>10</center>

a privilege log.

D.   Color. Consistent with the preservation objective stated in § I.D above, the parties will, wherever practicable, produce color images in color. If an original color image is produced in black and white, the receiving party may for good cause request the producing party to produce the image in color. Following a request for color production, the parties will meet and confer on the merits of the request and, if warranted, a reasonable means of providing the requested documents in color.

E.   Production Media. All document productions should have an alphanumeric volume name. If a production contains multiple volumes, each volume should be contained in an appropriately named folder at the root of the drive. Volumes should be numbered consecutively (ABC001, ABC002, etc.). Production media should be labeled with the name of this action, the identity of the producing party, and the following information: volume name, production range(s), and date of delivery. If the production range will not fit on the production media label, that information should be included in the cover letter accompanying the production. If the producing party encrypts the production, the producing party shall provide an explanation of how to decrypt the files.

F.   Replacement Documents. Any documents that are replaced in later productions shall be given the same, unique Bates number in addition to clearly designated as replaced, by appending an "R" to the production number prefix. When a party produces a replacement production indicated by "R," the receiving party must discard all copies, including working copies, of the original, replaced image.

G.   Exceptions. The producing party shall make reasonable efforts to limit the extent

of documents that cannot be processed. For documents and ESI that convert with errors to TIFF (e.g. oversized drawings, picture files, etc.), the parties will meet and confer regarding a reasonable alternative form of production.

III.   Miscellaneous Provisions

A.   Modifications. Any practice or procedure set forth herein may be varied by agreement of the parties and confirmed in writing. For purposes of this provision, an agreed upon variance may be evidenced by an e-mail confirmation from the party from whom allowance for such variance has been sought.

B.   Objections. Nothing in this Order shall be interpreted to require disclosure of information protected by the attorney-client privilege, work-product doctrine, or any other applicable privilege or immunity. The parties do not waive any objections as to the production, discoverability, admissibility, or confidentiality of documents or ESI.

C.   Rolling Productions. The parties will diligently produce responsive documents and information without undue delay. The parties anticipate producing large volumes of documents and productions will be made on a rolling basis. The parties will use their best efforts to substantially complete all document productions 45 days prior to the close of discovery as set forth in the Pretrial Scheduling Order (Document No. 91) or as further ordered by the Court.

D.   Inadvertent Disclosures. Inadvertent disclosures that a producing party claims to be privileged or protected by the attorney client privilege or attorney work product protection will be governed in accordance with the Confidentiality Order.

E.   Transparency of Process. The producing party shall maintain a record of the

12

material elements of its preservation, search and review methodologies, and any quality assurance activities undertaken to assess the accuracy and completeness of its preservation and production. Absent further order of the Court upon good cause shown and a determination that such disclosure does not violate any privilege of the producing party or otherwise prejudice that party, such record need not be disclosed to an adverse party, but a party may make such disclosure in the course of collaboration, subject to procedure outlined herein at Paragraph I.G. Material elements of search and review methodologies may include, but are not limited to, search terms used, search term occurrence analysis, the use of stop words, the construction of search queries, and the use of analytical algorithms, as may be applicable. Quality assurance activities include, but are not limited to, statistical sampling, reviewer overturn rates, training round stability, and calculations of precision and recall, as may be applicable.

F.   <u>Discoverability of Technology System</u>. Subject to the procedures set forth in the Federal Rules of Civil Procedure, the Court's Local Civil Rules, other orders of the Court, and a party's claim of privilege or objections due to burden or relevance, the parties may conduct discovery into the following subject matters:

1.   The architecture of a party's information technology infrastructure;

2.   The existence and scope of a party's information governance systems, data management systems, and information life-cycle policies; or

3.   The location and format of information likely to be discoverable.

G.   <u>Privilege Logs</u>. Each party shall provide a privilege log identifying all documents withheld in their entirety or produced but redacted on the basis of privilege and/or

work product claims in the discovery process that complies with the requirements of the Federal Rules of Civil Procedure. The withholding party's log shall be in Excel format or any other format that permits electronic sorting and searching. The privilege/redaction logs shall be detailed enough to enable other parties to assess the applicability of the privilege asserted; shall identify the author(s), attorney(s) and client, any and all recipient(s), [**PLAINTIFFS**: all persons or entities known to have been furnished the Document or informed of its substance], created date, beginning and ending Bates or document identification numbers, privilege and/or doctrine relied upon, and the subject matter of the document; and shall provide a description of the Document and/or Document family by type and content in conformity with the Federal Rules of Civil Procedure. For emails included in the privilege log that include attachment(s), the information above must also be provided for any such attachment(s). Each privilege/redaction log shall include the entries of all prior privilege logs. To the extent any previous designations are subsequently modified or withdrawn, they shall be so noted within 15 days of any decision to modify or withdraw any previous designations in an amended log to ensure clarity. The parties shall have no obligation to log privileged communications between outside counsel, in-house counsel, and their respective legal support staff and their respective client(s) generated on or after December 20, 2018. Privilege logs shall be provided to the receiving party within 30 days after each production of documents.

H.   Preservation. The Parties represent that they have taken reasonable measures to preserve documents (including electronic data) from its active information systems

14

that are potentially responsive to the claims made, the defenses asserted, and the

discovery propounded. The Parties further represent that their ongoing efforts are

designed to preserve relevant and potentially relevant documents and data in a

timely manner.

I.   Non-Party Subpoenas. A party that issues a non-party subpoena shall timely notify

other Parties when it receives non-party productions of documents or other

responses from the non-party, and shall provide copies of all such productions in

the format in which they were received from the non-party.

IT IS SO ORDERED

Dated: _____, 2020   _____
                                                UNITED STATES MAGISTRATE JUDGE

AGREED TO BY:

ANSA ASSUNCAO, LLP   **LITE DEPALMA GREENBERG, LLC**
James S. Coons   Bruce D. Greenberg
100 Matawan Road, Suite 410   570 Broad Street Suite 1201
Matawan, NJ 07747   Newark, NJ 07102

Attorneys for Defendants   *Attorneys for Plaintiffs*

15

Appendix 1: TIFF Image File Specifications

TIFF image files created for this litigation pursuant to this Order shall comport with the following specifications:

- The TIFF image file shall be a Group IV compression, black-and-white, single-page 1-bit TIFF image file using a 300 x 300 dots-per-inch (DPI) optical scan resolution and an 8.5 x 11 inch page size, and reflect the full and complete information contained in the original source Document, except as modified by any redactions. Documents that in the producing party's reasonable judgment require a different resolution or page size may be excepted from this requirement.

- Original document orientation should be maintained (*i.e.* an original landscape document should be produced in landscape format).

- Each TIFF image file shall be branded with a legible, unique Bates number in the lower right corner, positioned so as not to interfere with reading the document. The Bates number shall: (1) be unique across the entire document production; (2) be a combination of an alphabetic prefix along with an 8-digit number (*e.g.* ABC00012345), wherein the numeric portion shall be zero-padded leftwards as needed to comprise 8 digits, (3) contain no special characters or embedded spaces, and (4) be numerically sequential within a given document. If a production number or set of production numbers is skipped, the skipped number or set of numbers will be communicated to the receiving party.

- Confidentiality designations, if any, will be branded on the lower left corner of the applicable image, also positioned so as not to interfere with reading.

- The resulting TIFF image file shall be named according to the naming convention *number.TIF*, where "*number*" is the Bates number of the corresponding page. File names for these files shall not contain any special characters or embedded spaces.

- Image files shall be grouped in folders on the production media of no more than 1,000 TIFF files each unless necessary to prevent a file from splitting across folders. Files will not be split across folders, and separate folders will not be created for each file.

Appendix 2: Data Load File Specifications

The data load file is a delimited text file containing data about each document needed for use with standard litigation support software.

The file name of the data load file should contain the volume name of the production media, although additional description information may be provided after the volume name. For example, both "ABC001.dat" and "ABC001_metadata.dat" would be acceptable names for a data load file having a production volume named "ABC001." File names shall not contain any special characters or embedded spaces.

Unless other delimiters are specified, the data in each field should be separated using standard Concordance delimiters (Comma – ASCII character 20 (¶), Quote – ASCII character 254 (þ), Newline – ASCII character 174 (®)). A semicolon should be used as a multi-entry separator. A carriage-return line-feed should be used to indicate the start of the next record.

The first line of the data load file must contain each field name, separated by a delimiter, corresponding to the order in which the data appears in the file.

Each production has one index file, in .DAT file format. The format of the DAT file should use UTF-8 encoding.

Load files should not span across media (*e.g.* CDs, DVDs, hard drives, etc.); a separate volume should be created for each piece of media delivered.

Data for documents shall be produced in only one data load file throughout the productions, unless that document is noted as being a replacement document in the Replacement field of the data load file.

Metadata Fields to be Included in the Data Load File

For all documents or electronic files produced, the following metadata fields for each document or electronic file, if available at the time of collection and processing and unless such metadata fields are protected from disclosure by attorney-client privilege or work-product immunity or otherwise prohibited from disclosure by law or regulation, including the European Data Privacy Regulation, will be provided in the data load file except to the extent that a document or electronic file has been produced with redactions.

The term "Scanned Docs" refers to documents that are in paper form at the time of collection and have been scanned into *.tif images. The term "Email and E-Docs" refers to files that are in electronic form at the time of their collection.

823032.1

Field Name	Sample Data	Scanned Docs	Email and E-Docs	Comment
PRODBEG	ABC00000001	Yes	Yes	Beginning production number
PRODEND	ABC00000008	Yes	Yes	Ending production number
PRODBEGATT	ABC00000009	Yes	Yes	Beginning production number of parent in a family
PRODENDATT	ABC00001005	Yes	Yes	Ending production number of last page of the last attachment in a family

Field Name	Sample Data	Scanned Docs	Email and E-Docs	Comment
ALLCUSTODIAN	Smith, John	Yes	Yes	Custodian(s) that possessed the document or electronic file—multiple custodians separated by semicolon
FILEDESC	Microsoft Office 2007 Document	N/A	Yes	Description of the type file for the produced record.
FOLDER	\My Documents\Document1.doc	N/A	Yes	Original electronic source folder for the record produced (no data need be included if typical folder structure is not applicable to file).
FILENAME	Document1.doc	N/A	Yes	Name of original electronic file as collected.
DOCEXT	DOC	N/A	Yes	File extension for email or e- doc
PAGES	2	Yes	Yes	Number of pages in the produced document or electronic file.
AUTHOR	John Smith	N/A	Yes	Author information as derived from the properties of the document.
DATECREATED	10/09/2005	N/A	Yes	Date that non-email file was created as

				extracted from file system metadata
DATELASTMOD	10/09/2005	N/A	Yes	Date that non-email file was modified as extracted from file system metadata
SUBJECT	Changes to Access Database	N/A	Yes	"Subject" field extracted from email message or metadata properties of the document
FROM	JohnBeech@sample.com	N/A	Yes	"From" email address extracted from email message
TO	JaniceBirch@sample.com	N/A	Yes	"To" email address(es) extracted from email message
CC	FrankMaple@sample.com	N/A	Yes	"Cc" or "carbon copy" email address(es) extracted from email message
BCC	JohnOakwood@sample.com	N/A	Yes	"Bcc" or "blind carbon copy" email address(es) extracted from email message
DATESENT	10/10/2005	N/A	Yes	Sent date of email message (mm/dd/yyyy format)
TIMESENT	10:33 am	N/A	Yes	Sent time of email message, time zone set to GMT
DATERCVD	10/10/2005	N/A	Yes	Received date of email message (mm/dd/yyyy format)
TIMERCVD	10:33 am	N/A	Yes	Received time of email message, time zone set to GMT
CONFIDENTIALITY	CONFIDENTIAL	Yes	Yes	Text of confidentiality designation, if any
REDACTION	REDACTED	Yes	Yes	User-generated field indicating whether a document was redacted.
TEXTPATH	Text\001\001\ABC00000001.txt	Yes	Yes	Path to *.txt file containing extracted or OCR text

19

NATIVEPATH	Natives\001\001\ABC00000001.xlsx	N\A	Yes	Path to files supplied in native format.
MD5 HASH	30999747f4e6d7bef786e614ff2cf4b0	N/A	Yes	MD5 Hash for electronic document
REPLACEMENT	REPLACEMENT	Yes	Yes	"Replacement" indicates the image is a replacement for a previously produced image; otherwise blank.

Field Name	Sample Data	Scanned Docs	Email and E-Docs	Comment
PRODVOL	VOL001	Yes	Yes	Name of the Production Volume

20

Appendix 3: Text File Specifications

The text file will contain the text in the respective document, omitting any text redacted from the produced document.

The file naming convention for the text file shall be in the form number.TXT, where "number" is the Bates number for the first page of the respective document TIFF image. File names shall not contain any special characters or embedded spaces.

The full path and file name of the text file must be provided in the data load file as specified in Appendix 2.

OCR text for scanned images of paper documents should be generated by applying optical character recognition processes to the image of the document. The parties will endeavor to generate accurate OCR and will utilize quality OCR processes and technology. OCR shall be performed on a document level and provided in document-level text files. OCR text should not be delivered in the data load file or any other delimited file.

For electronic files, searchable text shall be extracted from the file and included in the text file. If the electronic file has no extractable text (*e.g.* a non-searchable PDF file), the document will be OCR'd, and file-level OCR text will be provided in lieu of extracted text. Extracted text shall likewise be produced in document-level text files. The same extraction or OCR process applied to other files shall be applied to hidden language, hidden fields or hidden metadata in an electronic document, excluding any redacted information, and the results reflected in the appropriate metadata fields produced.

The production of relevant ESI in searchable, full text format is limited to those forms of ESI that have text (in other words, any non-text formats [e.g., .wav and .jpeg] would not produce any corresponding text files). Each extracted text file will be named according to the first Bates number of the corresponding electronic document.

The extracted text shall also conform to the Unicode standard.

21

Appendix 4: Image Load File Specifications

The name of the image load file should mirror the name of the delivery volume, and should have an .OPT extension (*e.g.* ABC001.OPT). File names shall not contain any special characters or embedded spaces.

The volume names should be consecutive (*e.g.* ABC001, ABC002, etc.).

Every image in the delivery volume should be contained in the image load file, one row in the load file per TIFF image. The total number of documents referenced in the image load file should match the total number of designated documents in the data load file for that production.

Load files should not span across media (*e.g.* CDs, DVDs, Hard Drives, Etc.), *i.e.*, a separate volume should be created for each piece of media delivered.

22

823032.1

Appendix 5: Native File Specifications

The file naming convention for the native file shall be in the form number .EXT, where "number" is the Bates number for the respective placeholder TIFF image. File names shall not contain any special characters or embedded spaces.

The full path and file name of the native file must be provided in the data load file as specified in Appendix 2.

823032.1